Statistical Framework for Event Identification and Assessment of Seismic CTBT Monitoring Capability

Mark D. Fisk
Mission Research Corporation, Santa Barbara, CA
and
Henry L. Gray and Gary D. McCartor
Southern Methodist University, Dallas, TX

Contract No. F19628-93-C-0117 (Sponsored by ARPA)

ABSTRACT

This effort focuses on several main areas regarding event identification within the context of CTBT monitoring: (1) development and integration at the Center for Monitoring Research (CMR) of a software system for event characterization/identification; (2) assessment of regional seismic event identification performance; (3) analysis of GSETT-3 Alpha network data to assess the availability and utility of event characterization parameters for global monitoring; and (4) investigation of robust multi-sensor data fusion techniques for outlier detection and event classification.

We have been implementing a statistical event identification framework. The fundamental methods consist of statistical tests for outlier detection and event classification, as well as algorithms to test appropriate assumptions regarding the data in order to optimize and ensure validity of the results. Our methodology: (1) accurately treats statistical uncertainties of all discriminants used; (2) provides flexibility to incorporate any discriminant and assess its utility objectively; (3) can identify nuclear tests in regions for which relevant ground-truth training data may or may not exist; (4) can function in an automated mode to flag, categorize and/or rank anomalous events; (5) can control the false alarm rate; (6) and provides a rigorous and defensible framework with which to report and justify the results. To also serve as an interactive analysis tool, we have been developing an X Window graphical user interface, featuring a wide range of displays and exploratory data analysis tools. We have applied the methods to seismic events recorded by the ARCESS and GERESS arrays, station WMQ in China, and stations KNB and MNV in the western U.S. Results show that useful monitoring can be performed, currently down to magnitude 3, for regions that are well-covered by at least one seismic station or array. Overall, 91% of explosions were identified as outliers with a false alarm rate slightly higher than the target rate of 1%. In addition to the geological diversity of the regions, these results were obtained for events with a wide range of epicentral distances and magnitudes.

We have also compiled various statistics regarding 1,786 events detected by a set of 30 Alpha stations, currently transmitting data to the CMR, during GSETT-3. We discuss the availability and utility of event characterization parameters, including location (offshore vs. onshore), hypocentral depth and detected depth phases, Ms:mb, and high-frequency regional amplitude ratios.

We have also been addressing the problem of how to best utilize discriminant data from multiple sensors. In general, there are numerous types of seismic and non-seismic (e.g., hydroacoustic and infrasonic) measurements from multiple stations/sensors that can be used as discriminants, and it is not clear from the outset how to best combine this information in a multivariate test for outliers. We have investigated three approaches to this problem. We have derived expressions for the power of each of the three tests and performed a parametric study to determine conditions for which a particular test is favored over the others. We also compared the power of these three tests using actual seismic data from stations KNB and MNV.

Key Words: CTBT, Event Identification, GSETT-3, Alpha Network, Outlier Detection

Statistical Framework for Event Identification and Assessment of Seismic CTBT Monitoring Capability

OBJECTIVE

This effort focuses on several main areas regarding event identification within the context of CTBT monitoring: (1) development and integration at the Center for Monitoring Research (CMR) of a software system for event characterization/identification; (2) assessment of regional seismic event identification performance; (3) analysis of GSETT-3 Alpha network data to assess the availability and utility of event characterization parameters for global monitoring; and (4) investigation of robust multi-sensor data fusion techniques for outlier detection and event classification.

We have been developing a statistical framework to identify seismic events, based on multivariate discriminant analysis. Our approach first utilizes confidence intervals for depth and Ms-mb, and confidence ellipses for location (and eventually hydroacoustic data) to screen events with high confidence that are caused by natural seismicity. Due to current limitations in estimating depth, location and Ms-mb for events below mb ~4.5, analysis of remaining events is based, at least in part, on regional discriminants which include high-frequency Pn/Lg and Pn/Sn. Due to regional variations and the lack of calibration data for nuclear explosions in most regions, we developed a robust multivariate outlier procedure to characterize events relative to previous seismic activity. The outlier method, based on a generalization of the likelihood ratio, has sufficient generality to utilize any combination of teleseismic and regional (as well as non-seismic) discriminants from single or multiple stations. Fisk et al. (1993, 1994) describe the methodology in detail, as well as numerous applications to seismic data. For regions where we do have training data for nuclear explosions, we have also developed a classification procedure (Baek et al., 1995; Fisk et al., 1993).

Fisk et al. (1994) applied the outlier approach to identifying seismic events in diverse geological regions, recorded by the ARCESS and GERESS arrays in Norway and Germany, stations KNB and MNV in the western U.S., and station WMQ in China. Results show that useful monitoring can be performed with this approach, currently down to magnitude 3, for regions that are well-covered by at least one seismic station or array. Overall, 264 of 290 (91%) explosions were identified as outliers and there were 3 false alarms out of 158 earthquakes (1.9%), slightly higher than the target rate of 1%. In addition to the diversity of the regions, these results were obtained for events with a wide range of epicentral distances and magnitudes. Recently we have adapted our multivariate outlier approach to categorize events rather than providing a "yes/no" decision. Categories are intentionally defined in a non-judgemental manner, but with a rigorous probabilistic interpretation, by the degree to which an event differs from previous seismicity in a given region. This approach fuses multivariate data in a composite metric to focus on anomalous events.

RESEARCH ACCOMPLISHED

Analysis of GSETT-3 Data. Recently we have compiled various statistics regarding seismic events detected by a set of 30 Alpha stations currently transmitting data to the CMR, to assess the numbers and characteristics of events that can be expected to be observed by the Alpha network during a given period, including the availability and utility of event characterization parameters that can be used to identify them. Events analyzed include those which occurred between 11 January 1995 and 12 February 1995, during GSETT-3. Regional discriminants and M_S were computed for the events during this period. We examine how many events can be identified with high confidence as due to natural seismicity, based on teleseismic measures of depth, location and M_S:mb. Of the remaining events, we examine how many have useful regional P/S and P/Lg values such that they can potentially be identified. Last, we evaluate the number of remaining ambiguous events which lack adequate discriminant data to identify them.

Figure 1 shows locations of 30 Alpha stations that are currently transmitting data to the CMR, as well as epicentral locations of 1,786 events, which occurred between 11 January 1995 and 12 February 1995 and were reviewed by seismic analysts. Origin, association and event characterization data for these events were retrieved from the *origin*, *origerr*, *assoc* and *originamp* tables in the *IDCWDB* database at the CMR. Of the 1,786 events during this period, there were a total of 634 events within regional distance ($\Delta < 20^o$) of at least one of the 30 Alpha station.

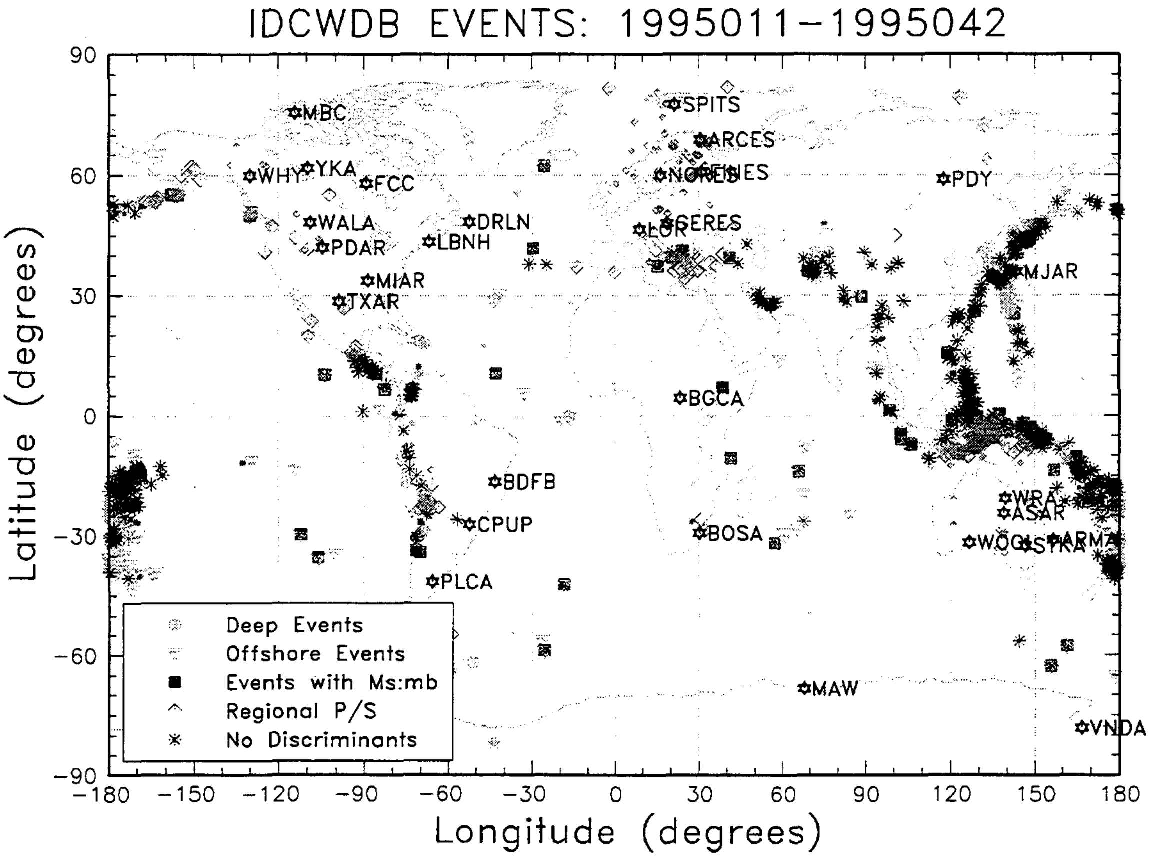

Figure 1. Locations of 30 Alpha stations and 1,786 seismic events recorded between 1995011 and 1995042. Marker types indicate events with various available discriminants or none.

Location Estimates and Uncertainty. Figure 2 shows the distribution of events versus the area of their 90% confidence location ellipses. Roughly half of the events have 90% location ellipses with areas greater than 10,000 km^2. Future use of Beta and Gamma station data, as well as data from future Alpha stations, should help to reduce the size of the location error ellipses. Comparisons by North (1995) of event locations in Canada, based on Alpha station data, to Canadian NDC locations using Beta and Gamma station data indicate that a significant percentage (11/16) of the 90% confidence location ellipses do not contain the Canadian NDC locations, accounting for their uncertainties as well. This may indicate that there are significant biases in location estimates which are not treated in the location uncertainty analysis. Note that of the 1,786 events during this period, 609 have 90% confidence location ellipses entirely offshore.

Depth and Detected Depth Phases. Figure 3 shows the cumulative distribution of events, with and without at least one detected depth phase, versus the upper bound of the 95% depth confidence interval. There were 292 events with 95% depth confidence intervals deeper than 10 km, of which there were 58 with at least one detected depth phase. Thus, using a 95% depth confidence interval, 16% of the events can be classified as deep natural events. If we also require at least one detected depth phase, only 3% of the events can be classified as deep natural events with high confidence.

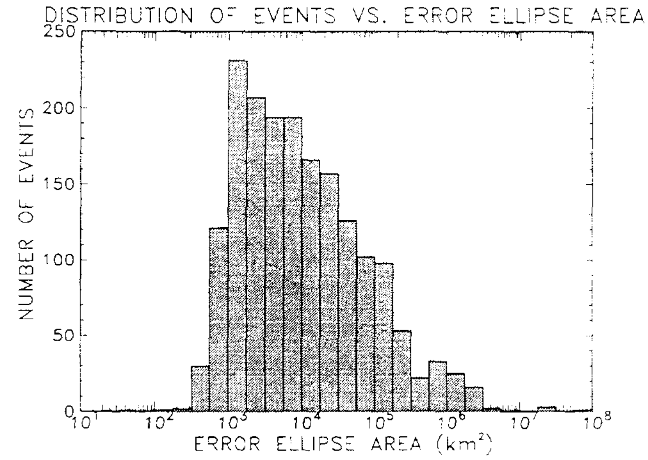

Figure 2. **Distribution of events versus 90% confidence location ellipse area.**

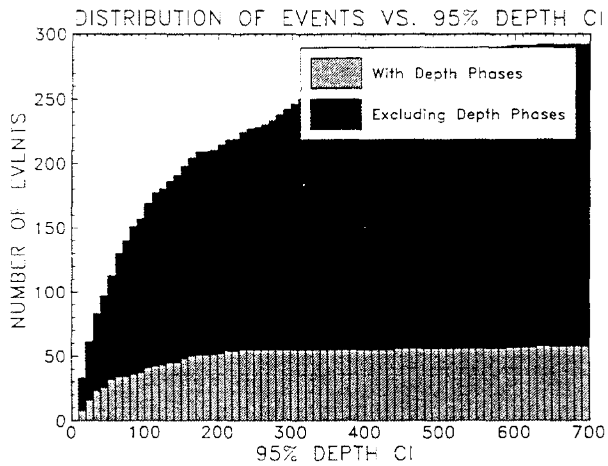

Figure 3. **Cumulative distribution versus upper bound of 95% depth confidence interval.**

Event Magnitudes and Ms:mb. Of the 1,786 events, there were 1,459, 143 and 398 events for which mb, M_S and M_L, respectively, were measured. Of the events for which mb was measured, all but 35 (roughly 2%) were above mb 3. Comparisons of IDC mb values to those from the QED (Quick Epicentre Determination) of the U.S.G.S. NEIC found that the IDC mb values are about 0.3 units smaller than the QED magnitudes (IDC Performance Report, 3 February 1995). This suggests that roughly 98% of the events are above mb 3.3 and, thus, there are very few mining blasts to contend with for now. Figure 4 (left) shows a scatter plot of M_S versus mb for 143 events for which both mb and M_S were measured. All but 3 of these 143 events fall above the line defined by mb–M_S = 1.2. A similar plot of M_L versus mb is shown on the right of Figure 4. The scatter about the linear trend is greater than one magnitude unit. This has serious implications regarding the usefulness of M_L as an accurate or consistent measure of source size, and will likely require regional calibration.

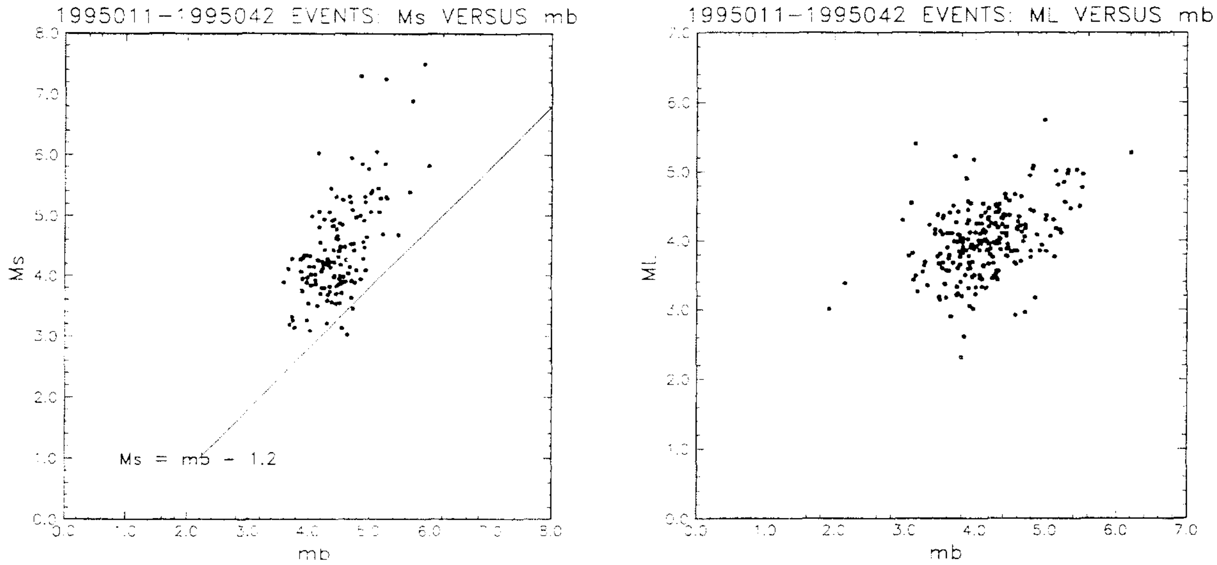

Figure 4. Scatter plots of M_S versus mb (left) and M_L versus mb (right).

High-Frequency Regional Amplitudes. Regional phase amplitudes were computed by automated software at the CMR. Absolute maximum amplitudes were measured on 2-4, 4-6, 6-8 and 8-10 Hz rms beams within predicted group velocity or time windows for Pn, Pg, Sn and Lg. Similar measurements were made of noise amplitudes in predicted pre-Pn, pre-Pg, pre-Sn and pre-Lg windows. The windows for predicted arrivals are used for all regions and stations. Figure 5 shows the number of regional events associated by each Alpha station, those for which Pn/Sn, Pn/Lg, Pg/Sn and/or Pg/Lg were computed, and the subset with SNR greater than 2 for Pn or Pg and Sn or Lg.

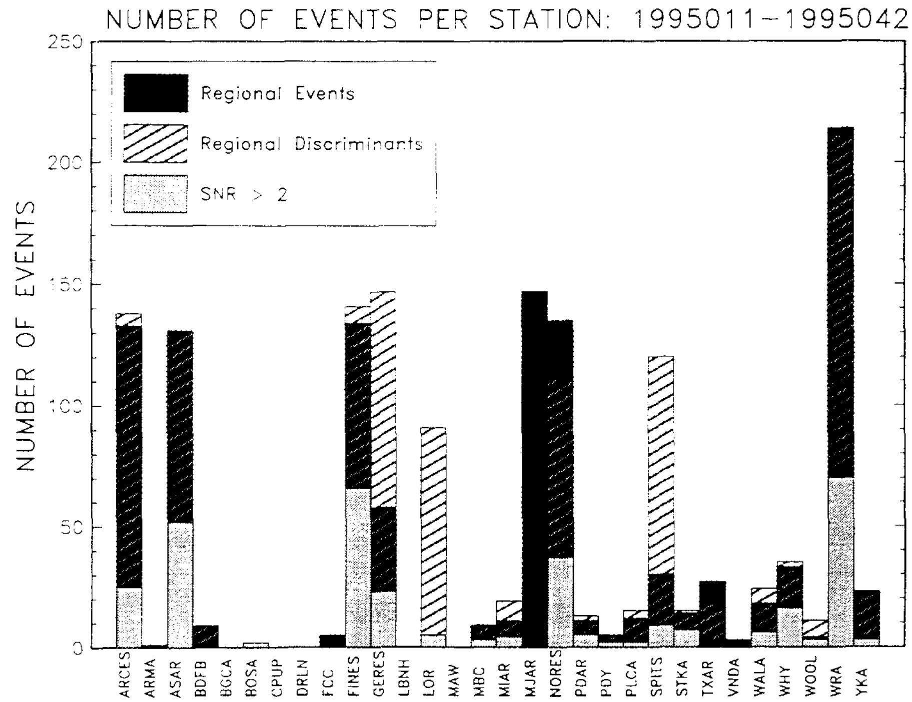

Figure 5. Number of regional events, and those with regional amplitude ratios, per Alpha station.

Except for MJAR, regional phase amplitudes were computed for most (487 of 634) regional events. For some stations, the number of events for which regional phase amplitudes were computed (given in the *originamp* table) exceeds the number of associated regional events in the *assoc* table. Only 249 of the 487 events had SNR greater than 2 for at least one regional P phase and Sn or Lg. Thus, less than 40% of the events detected within regional distances have useful high-frequency regional amplitude ratios. Figure 6 shows plots of Pn/Sn in four frequency bands (2-4, 4-6, 6-8 and 8-10 Hz) for each Alpha station. Similar plots (not shown here) exist for Pn/Lg, Pg/Sn and Pg/Lg. The markers are coded by SNR: circles indicate SNR below 2 for either numerator or denominator, while asterisks indicate SNR greater than 2 for both. Note that the majority of amplitude ratios do not satisfy a criteria of SNR > 2 for both numerator and denominator. Examination of the amplitude ratios also indicates that there are a very wide range in values (e.g., from 10^{-4} to 10^{+4} at station WRA) for events which are predominantly earthquakes.

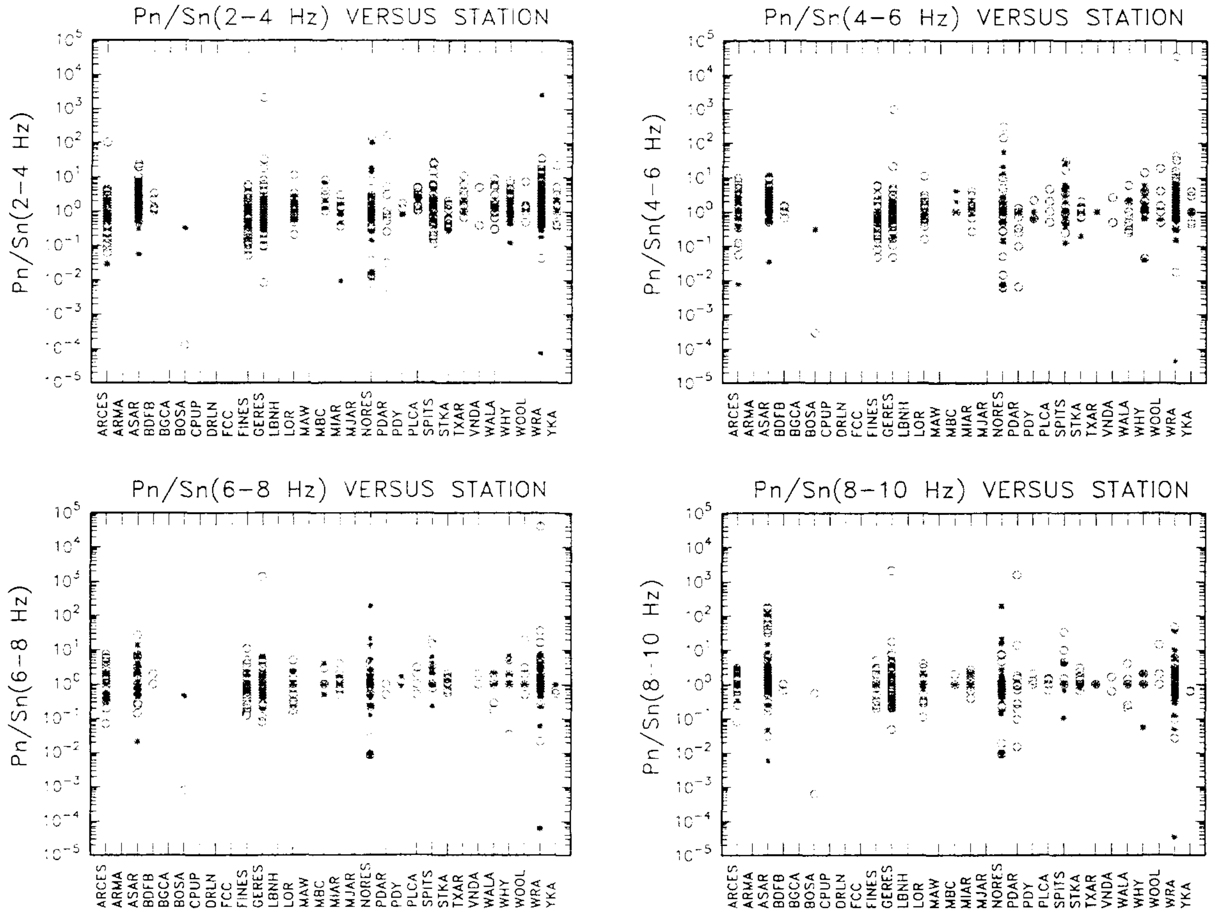

Figure 6. Pn/Sn in 4 frequency bands, plotted by station. Circles indicate SNR < 2, while asterisks indicate SNR > 2.

By studying 12 events with peculiarly high amplitude ratios (e.g., greater than 100), F. Ryall (1995) found the following discrepancies. There are cases for which the predicted phase windows do not correspond to actual phase arrivals or pre-phase noise, leading to anomalous regional phase amplitude measurements. For example, Pn times in the *originamp* table are about 60 sec later than P times in the REB (Reviewed Event Bulletin) for all 12 events except those in N. Europe, where they are about 30 sec late. As a result, the "Pn" amplitudes in the *originamp* table were measured well into the actual P coda, as were the "pre-Pn noise" amplitudes. Sn times in the *originamp* table are also about 30 sec late at the Australian arrays, while 15 and 5 sec late for two N. European events for which the REB had readings. The automated program measured amplitudes for crustal phases (e.g., "Pg") for events which were deep (e.g., > 190 km) and for which the REB arrivals were P and S. Amplitudes were also calculated for phases that were not associated (i.e., not listed in the REB *assoc* table). Theses discrepancies must be examined more thoroughly and resolved if these measurements are to have valid application to regional event characterization.

Number of Ambiguous Events. Overall, there were 883 events which can be identified by teleseismic measures of depth, offshore location (plus hydroacoustic data eventually) and/or M_S:mb. An additional 166 events have regional discriminants with adequate SNR, although further quality control is needed to validate these measurements. At best there are still 511 teleseismic and 226 regional events which <u>cannot</u> be identified based on available information considered. Figure 7 shows locations and 90% error ellipses of the 511 teleseismic events (asterisks) which cannot be classified as deep or offshore, do not have M_S:mb, and do not have any regional P/S or P/Lg discriminants. The diamonds indicate locations of the 226 regional events which do have at least one regional P/S or P/Lg discriminant, but with SNR less than 2 for either numerator or denominator (or both). Note that many of these events are at far regional to teleseismic distances from existing Alpha stations, but would be well within regional distances of future Alpha stations.

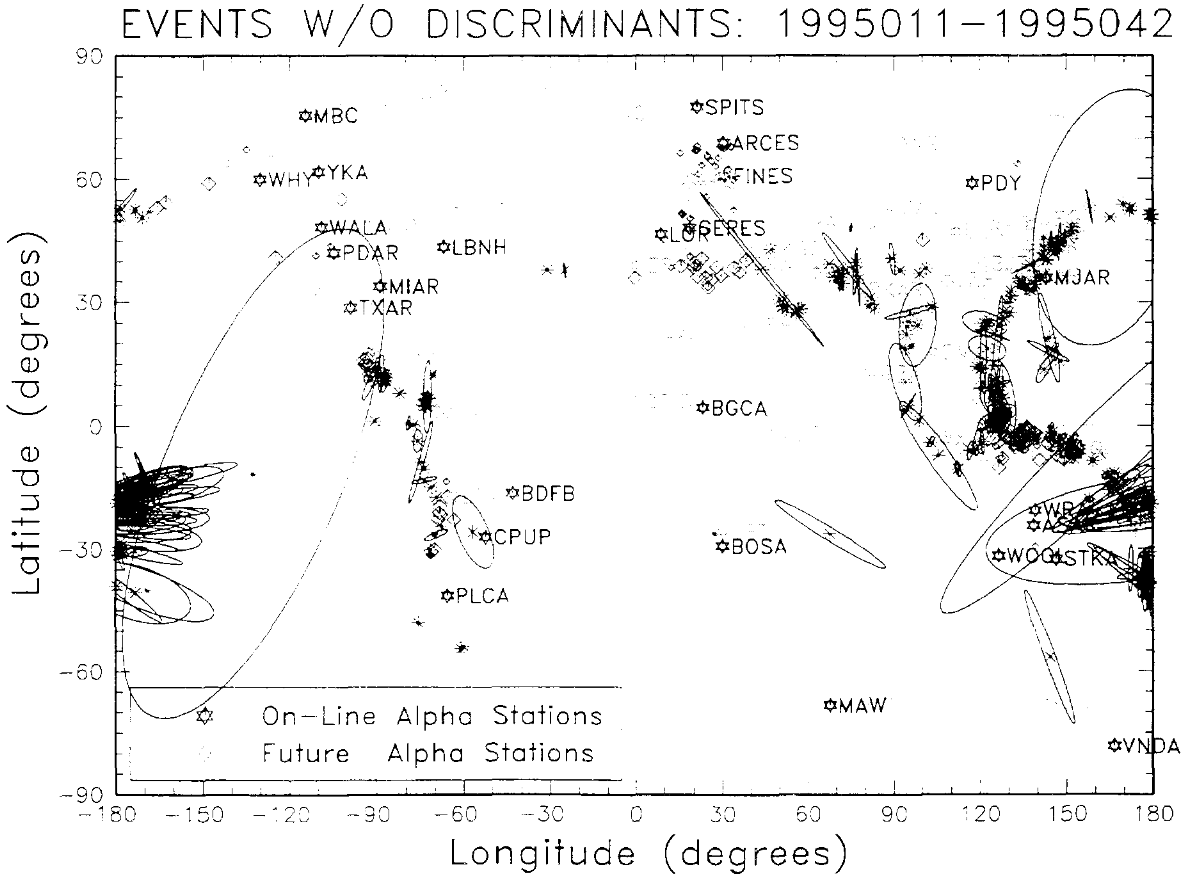

Figure 7. Locations of events for which none of the discriminants are available or have adequate SNR.

There are, however, a significant number of events within regional distances of existing Alpha stations (e.g., from MJAR in Japan) for which no regional phases were associated, even after analyst review. For events which occurred between 1995001 to 1995027, there were only 158 of 444 events within regional distances for which Pn or Pg and Sn or Lg were among the associated phases. Possible explanations for this include (1) regional phase attenuation for events at far regional distances; (2) regional phase blockage (e.g., for events southeast of GERES); (3) deep regional events (e.g., near MJAR) which do not produce crustal phases but whose 95% depth confidence intervals were not deeper than 10 km; and (4) possible phase association errors by the automated system which were not corrected after analyst review. Examples supporting the first three explanations have been found. Study of relevant waveforms is needed to determine whether an analyst could identify regional phases which were possibly missed by the automated system.

Last, Figure 7 shows epicentral location estimates of at least five events which appear to be clearly offshore, e.g., two in the north Atlantic Ocean, one in the south Pacific Ocean, and two in the Indian Ocean (one south and one west of Australia). The two in the north Atlantic are actually in the Azores Island region, while the two in the Indian Ocean and the one in the south Pacific have 90% error ellipses which overlap either Antarctica, Madagascar or North America.

Optimizing Multivariate Evidence for Outlier Detection. We have also been addressing a fundamental problem of how to best utilize multivariate discriminant data from multiple sensors. In general, there are numerous types of teleseismic and regional, as well as non-seismic (e.g., hydroacoustic and infrasonic) measurements from multiple stations/sensors that can be used as discriminants and it is not clear from the outset how to best combine this information in a test for outliers. We have investigated three approaches to this problem. The first (Test 1) is to insert all available features that are thought to discriminate into the outlier likelihood ratio (e.g., Fisk et al., 1993). The second (Test 2) is to perform an optimal weighting of some or all of the discriminants first and then insert the weighted combination(s) into the likelihood ratio. The third (Test 3) is to perform separate tests, based on the likelihood ratio and subsets of discriminants (e.g., for individual sensors), calling an event an outlier if it is found to be an outlier by any of the individual tests. The significance levels of the individual tests in this case must be appropriately modified in order to maintain the overall false alarm rate of the combined results.

We derived expressions for the power of each of these three tests for fixed significance level (i.e., fixed false alarm rate). The power is the probability of identifying an explosion as an outlier. To illustrate how the powers compare for a real situation, we used Pn/Lg measurements in the 6-8 Hz band at stations KNB and MNV for 35 earthquakes and 69 nuclear explosions on the Nevada Test Site. In this case, the estimates of the correlation, standard deviations and differences in the means of the Pn/Lg values for earthquakes and explosions are $\hat{\rho} = 0.228$, $\hat{\sigma}_1 = 0.485$, $\hat{\sigma}_2 = 0.416 = 0.856\hat{\sigma}_1$, $\Delta\hat{\mu}_1 = 1.104 = 2.276\hat{\sigma}_1$, and $\Delta\hat{\mu}_2 = 1.662 = 3.427\hat{\sigma}_1$. Under the assumption of normality (which was shown to be valid) and setting the significance level at $\alpha = 0.01$, the powers of Tests 1, 2 and 3 are 0.922, 0.954 and 0.926, respectively. Test 2 has the greatest power in this case, although all three tests have comparable power. For comparison, the powers of the tests at 0.01 significance level using either one of the stations is 0.363 for KNB and 0.938 for MNV. Thus, combining data from the two stations has far greater power than only using the worst station, comparable or greater power than using only the best station, and does not require that we know which station provides greater power which, in practice, we will not know in most situations.

We also performed a parametric study to assess more general conditions for which a particular test is favored over the others. Figure 8 illustrates results of this parametric power comparison. Shown are the differences of the powers of Tests 2 and 1 (left) and Tests 3 and 1 (right) versus correlation (ρ) for $\sigma_1 = \sigma_2 = 1$ (upper) and $\sigma_1 = 1$ and $\sigma_2 = 2$ (lower). Each curve corresponds to various values of $\Delta\mu_1$ and $\Delta\mu_2$ from 0 to 4. The results show that while there are regions of the parameter space (depending primarily on the correlations and separations of the discriminant means for different types of events) for which Test 2 or Test 3 can have the greatest power (e.g., by 5% to 10%), there are many cases for which they can provide dramatically poorer power than Test 1 (e.g., by as much as 90%). Test 1 generally has comparable or greater power than the other two tests over all parameters and there are no situations in which using Test 1 yields dramatically poorer results. Although we will eventually have considerable information regarding discriminant means and the covariance matrix for earthquakes, we will not know the discriminant means for nuclear explosions in most regions. Thus, the most robust procedure is to use Test 1 in which the full vector of discriminants is inserted into the likelihood ratio. This illustrates how fusion of multi-sensor data within a single test, which treats their correlations, provides the most robust results.

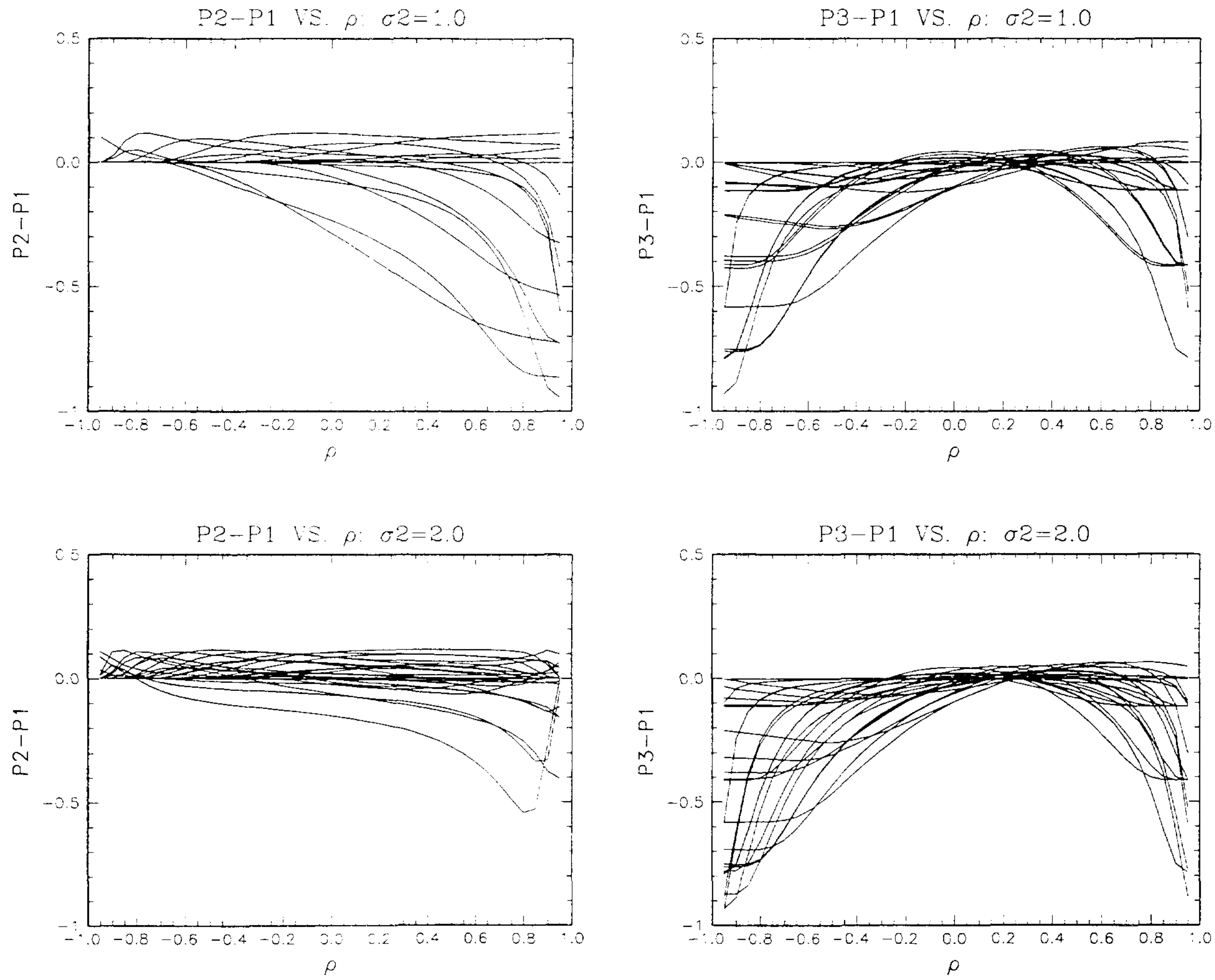

Figure 8. **Difference in the power of Tests 2 and 1 (left) and Tests 3 and 1 (right) versus correlation for various values of $\Delta\mu_1$, $\Delta\mu_2$ from 0 to 4, and for $\sigma_1 = \sigma_2 = 1$ (upper) and $\sigma_1 = 1$ and $\sigma_2 = 2$ (lower).**

CONCLUSIONS AND RECOMMENDATIONS

Preliminary results show that useful seismic monitoring can be performed with the outlier-detection approach, currently down to magnitude 3, for regions that are well-covered by at least one seismic station or array and provided useful discriminants are available. Results were obtained for diverse geological regions and a wide range of epicentral distances and magnitudes. In our outlier approach, we assume that events above mb 3 are predominantly earthquakes with potentially one or at least a small number of nuclear explosions. That is, we assume that there will be very few chemical mining blast above mb 3. Note that of the 1,459 events between 1995011 and 1995042 with mb measurements, there were only 35 below mb 3. Accounting for the bias of 0.3 units in the IDC mb values relative to QED mb values implies that roughly 98% of the events during this period were actually above mb 3.3. This suggests that the majority of events will fall in the regime where the assumption for application of the outlier test is valid.

Further analysis of GSETT-3 data, however, indicates that there are a significant number of events which cannot currently be processed to identify them based on the incomplete Alpha network and a current lack of sufficient event characterization processing capabilities.

First, requiring that the 95% confidence interval for depth be deeper than 10 km, 292 of the 1,786 events (16%) are classified as deep natural events. If we also require that at least one depth phase be detected (since there may be potential biases in location and depth estimates), only 58 events are classified as deep. Thus, it is important to resolve the issue of biases in location and depth estimates, and to assess whether improvements in associating depth phases can be made.

Second, there are 143 events (8%) with M_S:mb measurements. This relatively low percentage is due to the fact M_S measurements are currently made only for long-period instruments which are limited in number. Future work by S-Cubed includes array processing for broadband instruments, which should lead to a much higher percentage of events with M_S measurements. All but 3 of the 143 events satisfy M_S–mb > 1.2; however, the effect of magnitude biases has yet to be resolved.

Third, 609 events have 90% location error ellipses entirely offshore, but there were no hydro-acoustic (HA) data available to verify that they were not offshore explosions. Future plans by ARPA include collection of HA data and implementation of algorithms to compute relevant event characterization parameters. Incorporating seismic data from Beta stations could lead to smaller error ellipses, thereby improving the capability to classify offshore events with high confidence.

Fourth, there were 487 events for which regional phase amplitudes were computed. Of these, 249 have at least one Pn or Pg phase and one Sn or Lg phase with SNR greater than 2. We found, however, several discrepancies in these measurements, namely, some very large amplitude ratios (e.g., greater than 100) for events which are presumed earthquakes, a peculiarly high number of events with identical regional P phase and Sn or Lg amplitudes in the three highest frequency bands, and some events with amplitudes greater than 1000 but with SNR less than 2. We also found that the theoretical windows used in computing the regional phase and noise amplitudes often do not coincide with the true expected arrivals. These discrepancies must be resolved before these regional discriminants can be used with any degree of confidence.

Overall, there were 883 events which can be identified by teleseismic measures of depth, offshore location and/or M_S:mb. An additional 166 events have regional discriminants with adequate SNR, although further quality control is needed to validate these measurements. At best there are still 511 teleseismic and 226 regional events which cannot be identified based on available information considered. That is, 41% of the events during the period 1955011-1995042 cannot be classified as deep or offshore, and do not have M_S:mb or regional amplitude ratios with adequate SNR.

This situation should improve significantly as the Alpha network nearly doubles and as processing algorithms are integrated. Our future work focuses on integration at the prototype IDC of algorithms, developed by MRC and other ARPA contractors, to compute new event characterization parameters. We will also be implementing our event characterization/identification subsystem to combine this information optimally. To monitor below mb 3, much of the effort by MRC and other ARPA contractors is directed at implementing reliable automated methods to identify mining blasts. We also plan to extend our outlier approach to test potential nuclear explosions as outliers of a mixture of earthquake, mining blast and rock burst groups. In addition to implementing software for event characterization, our future work will also involve evaluation of monitoring capabilities as the prototype IMS infrastructure evolves.

While our statistical framework has been applied in the past to only seismic discriminants, it possesses the generality and flexibility to incorporate hydroacoustic, infrasonic and other forms of useful discriminant data. Lessons learned from our multivariate outlier analysis indicate that the most robust approach may be to fuse multi-sensor data directly, thereby treating correlations, rather than performing separate tests and then combining the results.

REFERENCES

Baek, J., H.L. Gray, and W.A. Woodward, J. Miller and M.D. Fisk (1995). A Bootstrap Generalized Likelihood Ratio Test in Discriminant Analysis, in press, *Comp. Stat. Data Anal.*

Fisk, M.D., H.L. Gray and G.D. McCartor (1994). Preliminary Assessment of CTBT/NPT Monitoring Capability, PL-TR-94-2300, Phillips Laboratory, Hanscom AFB, MA. ADA293188

Fisk, M.D., H.L. Gray and G.D. McCartor (1993). Applications of Generalized Likelihood Ratio Tests to Seismic Event Identification, PL-TR-93-2221, Phillips Laboratory, Hanscom AFB, MA, ADA279479.

IDC Performance Report - January 15-28, 1995 (February 3, 1995). Available through the Center for Monitoring Research, Arlington, VA.

North, R. (1995). GSETT-3 - Experience and Lessons Learned So Far, paper in *Proceedings of the CTBT Monitoring Technologies Conference*, 15-18 May 1995, Chantilly, VA, Sponsored by ARPA/NMRO, Arlington, VA.

Ryall, F. (1995). Private Communication.